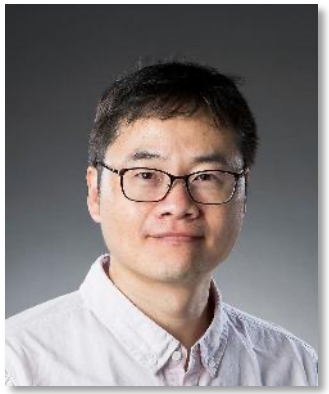
ICML
International Conference
On Machine Learning

MLV

HInT: Hypergraph Infusion at the Structural Layers Improves Table Understanding

Wonjin Lee

Soomi Jeong



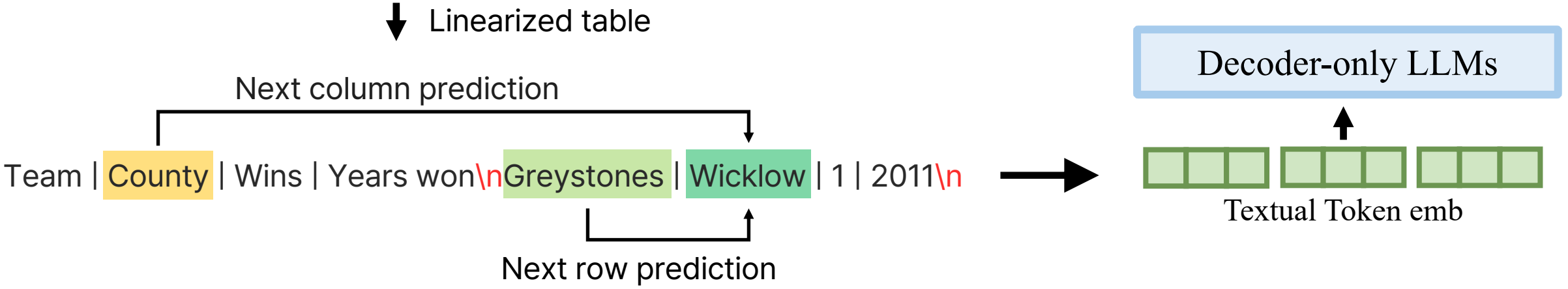
Kwang In Kim

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Key challenges: Loss of structural information

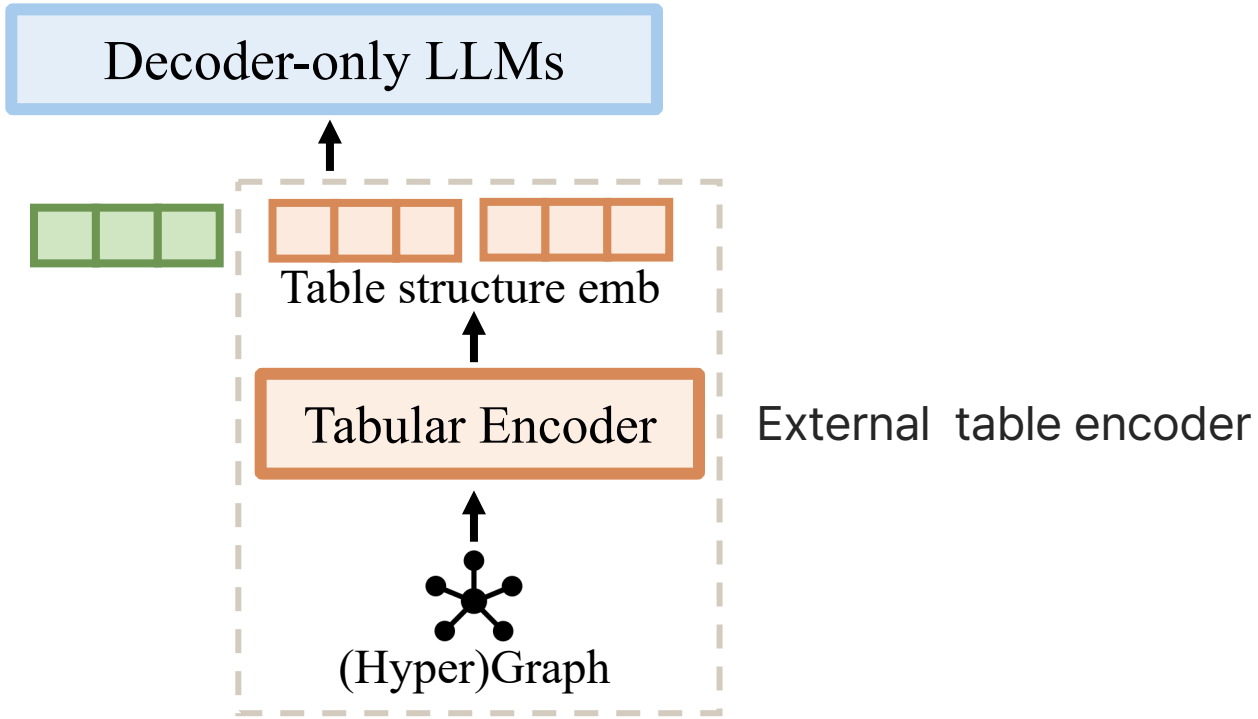
Linearizing tables obscures row/column relations and higher-order structure

Team	County	Wins	Years won
Greystones	Wicklow	1	2011
Ballymore Eustace	Kildare	1	2010
Maynooth	Kildare	1	2009



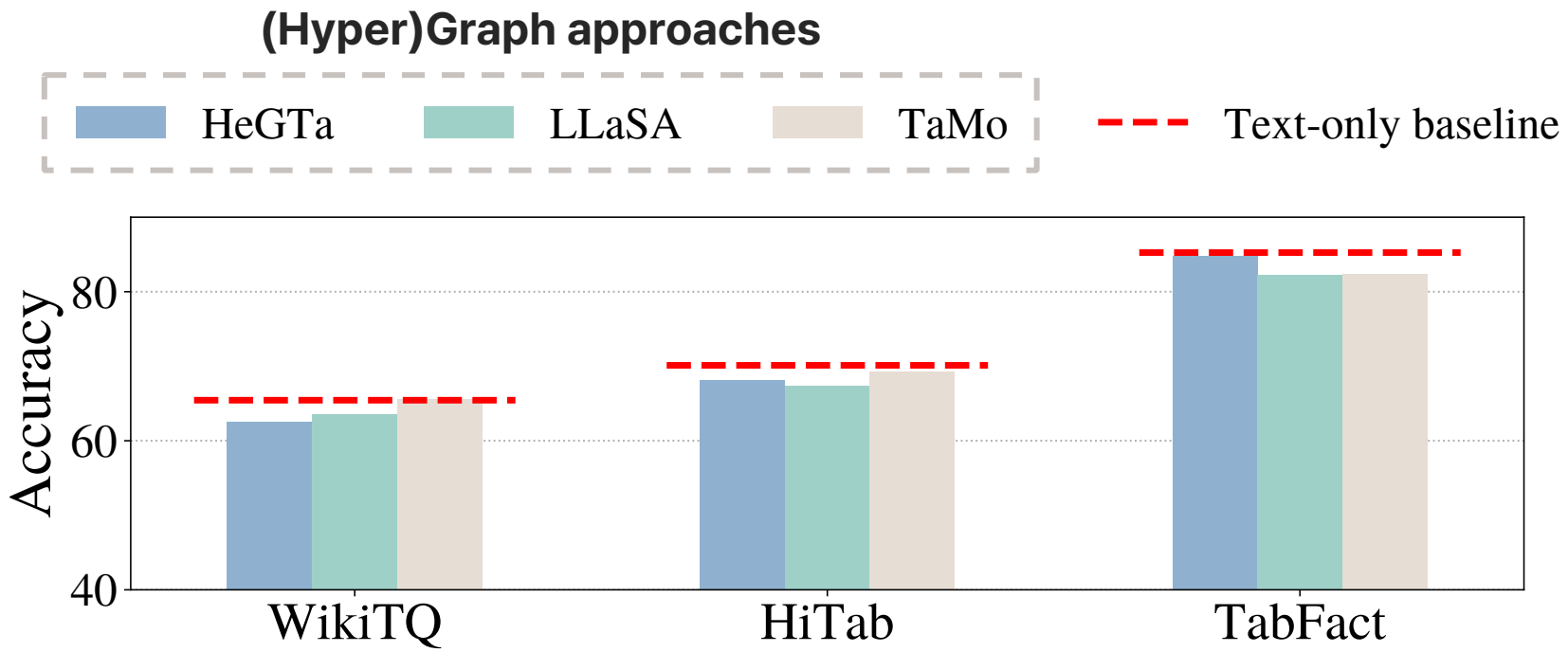
Prior works: Table encoder for structure modeling

Table structure is encoded separately and injected as structure-aware features



Prior works: Limitations of external table encoders

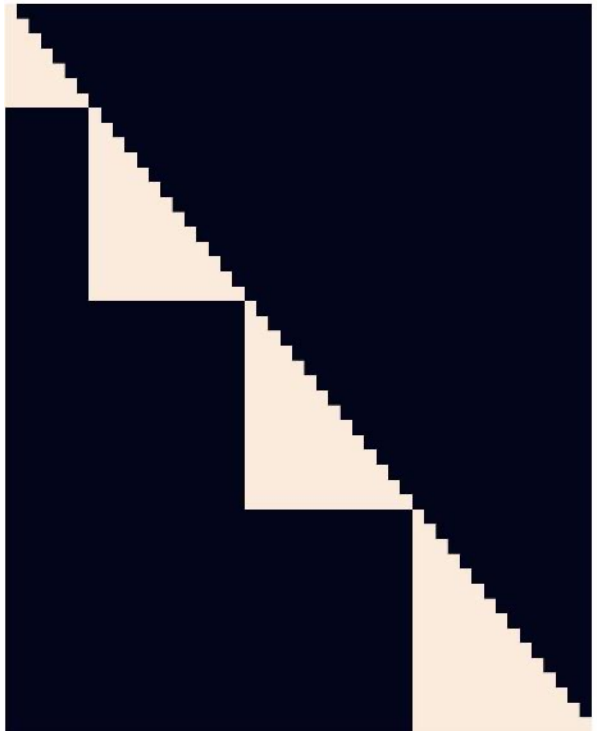
(Hyper)Graph approaches yields comparable, and in some cases lower performance than text-only baselines under multi-step reasoning setting



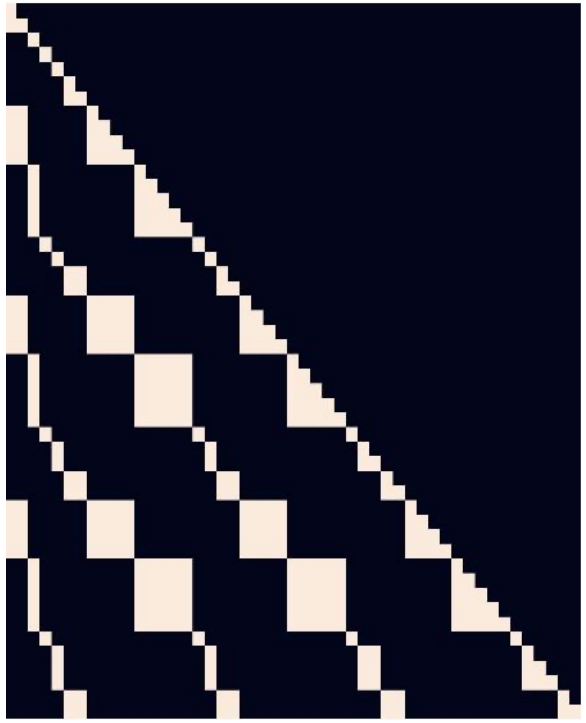
Key message: The bottleneck is not just extracting structure, but integrating it where reasoning happens inside the LLM

Our approach: Identifying structurally significant attention heads

Table structure is encoded as row-wise and column-wise attention masks



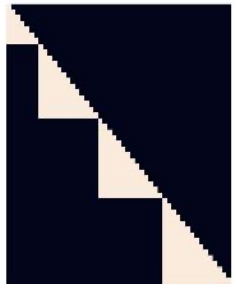
Row mask M^R



Column mask M^C

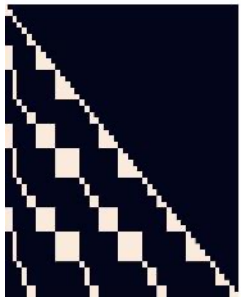
Our approach: Identifying structurally significant attention heads

We measure the alignment between attention patterns and the structural masks



Row mask $\mathbf{M}^R$

$$= w_0^0 \mathbf{A}_0^0 + w_1^0 \mathbf{A}_1^0 + \ldots + w_h^l \mathbf{A}_h^l$$



Column mask $\mathbf{M}^C$

$$= w_0^0 \mathbf{A}_0^0 + w_1^0 \mathbf{A}_1^0 + \ldots + w_h^l \mathbf{A}_h^l$$

$$\{\mathrm{A}_h^l | l = 1, \ldots, L, h = 1, \ldots, H\}$$

Our approach: Identifying structurally significant attention heads

We measure the alignment between
attention patterns and the structural masks

$$\mathcal{L}(\mathbf{w}) = \left\| \mathbf{m}^{\mathrm{R}} - \sum_{l,h} w_h^l \mathbf{a}_h^l \right\|_2^2 + \lambda \|\mathbf{w}\|_1,$$

$$\mathbf{a}_h^l = \mathrm{vec}(A_h^l), \qquad \mathbf{m}^{\mathrm{R}} = \mathrm{vec}(M^{\mathrm{R}}), \qquad \mathbf{w} = \{w_h^l\}_{l,h}.$$

$$\{A_h^l | l = 1, \ldots, L, h = 1, \ldots, H\}$$

We ranked the attention heads by the learned LASSO coefficients

Our approach: Identifying structurally significant attention heads

Removing the top ranked heads results in a larger performance drop than removing randomly selected heads

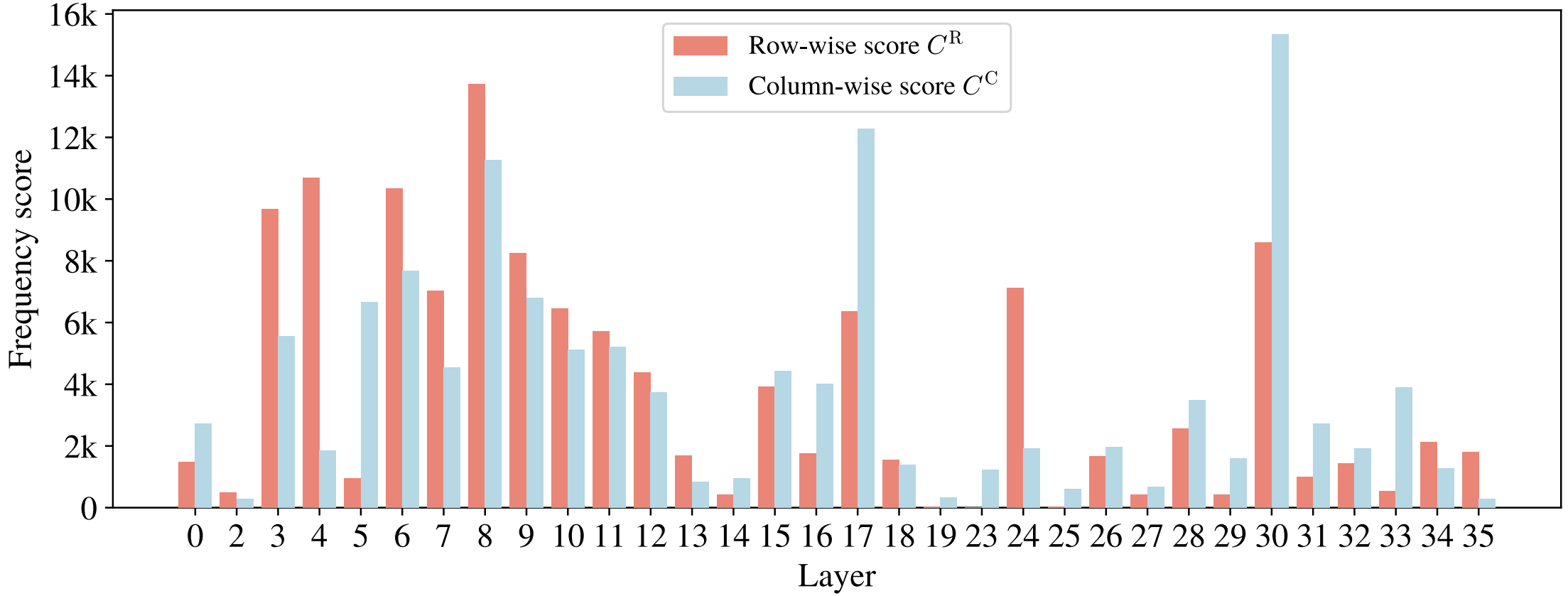
	Original	**Ablate structural heads**	Ablate random heads
Qwen 2.5-3B	47.51	**12.73**	32.29

(in % exact match; EM)

Our approach: Identifying structural layers

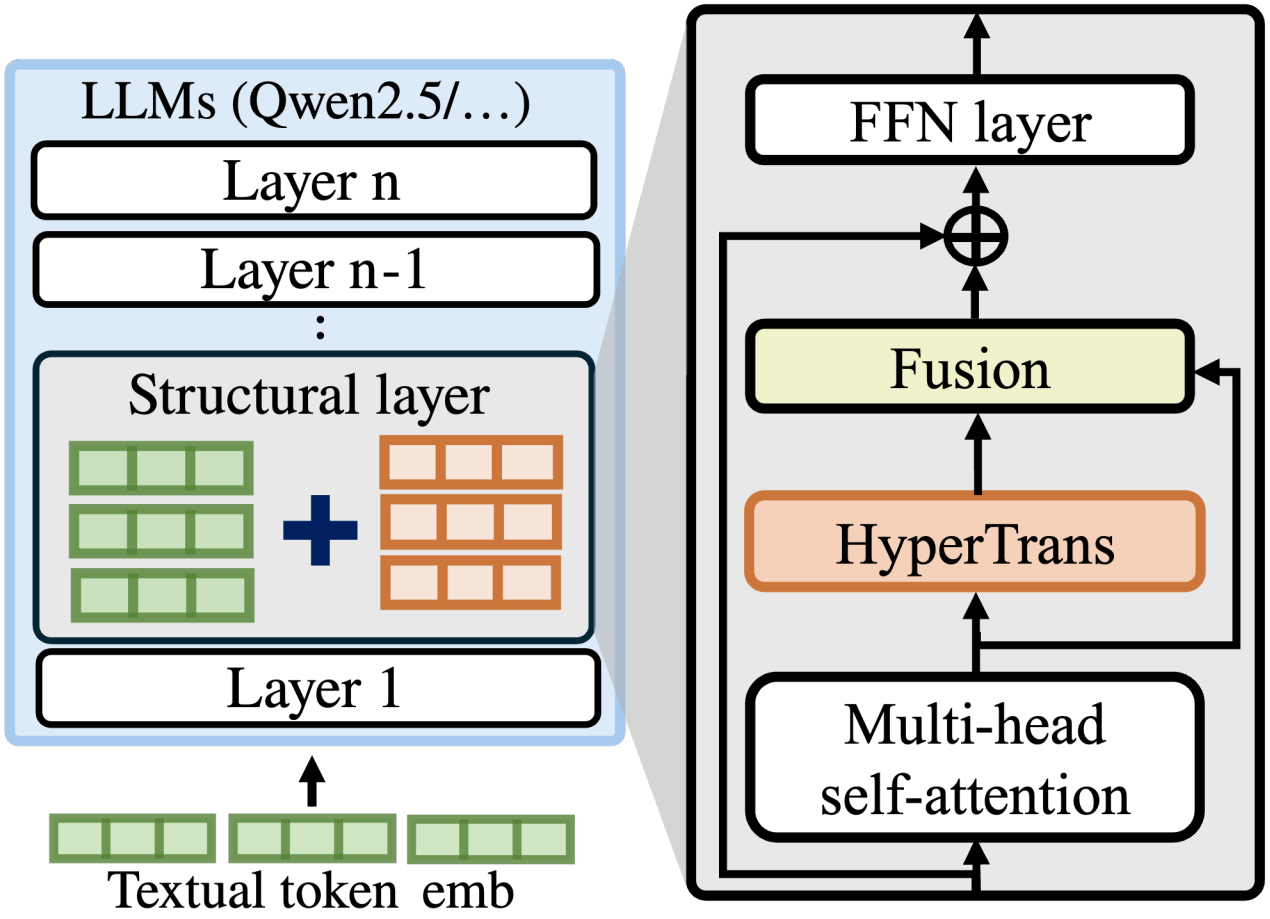
Structural layers (top-12)

$$\mathcal{L}^{\star} = \{3, 4, 5, 6, 8, 9, 10, 12, 15, 17, 24, 30\}$$



Our approach: Injecting structural information at structural layers

Converting token representations into node and hyperedge features



Our approach: Injecting structural information at structural layers

Given a table, we construct a hypergraph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

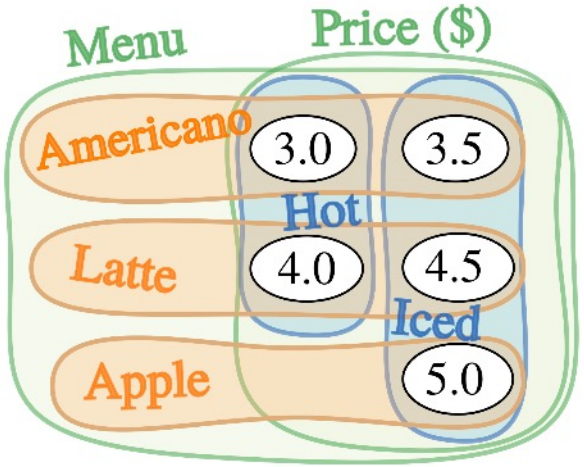
$\mathcal{V} = \{1, \ldots, N_c\}$: table cells

$\mathcal{E} = \{e_1, \ldots, e_{N_h}\}$: hyperedges

$$\mathcal{E} = \{\text{Menu},\ \text{Price (\$)},\ \text{Americano},\ \text{Latte},\ \text{Apple},\ \text{Hot},\ \text{Iced}\}$$

Menu	Price ($)	
	Hot	Iced
Americano	3.0	3.5
Latte	4.0	4.5
Apple		5.0

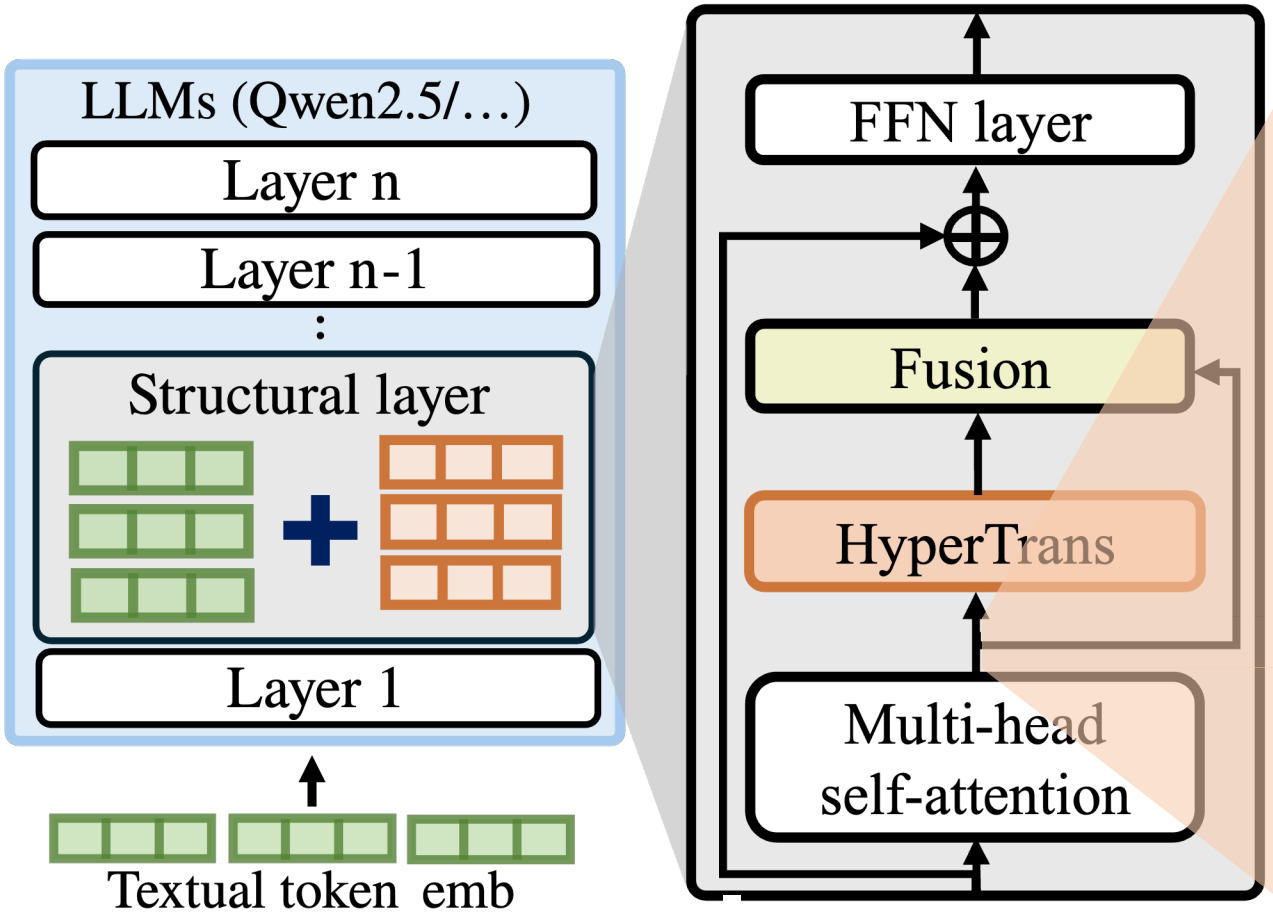
Table



Hypergraph

Our approach: Injecting structural information at structural layers

Converting token representations into node and hyperedge features



Node

$$\mathbf{v}_u = \frac{1}{|\mathcal{T}_u^{\mathcal{V}}|} \sum_{i \in \mathcal{T}_u^{\mathcal{V}}} \mathbf{u}_i$$

Hyperedge

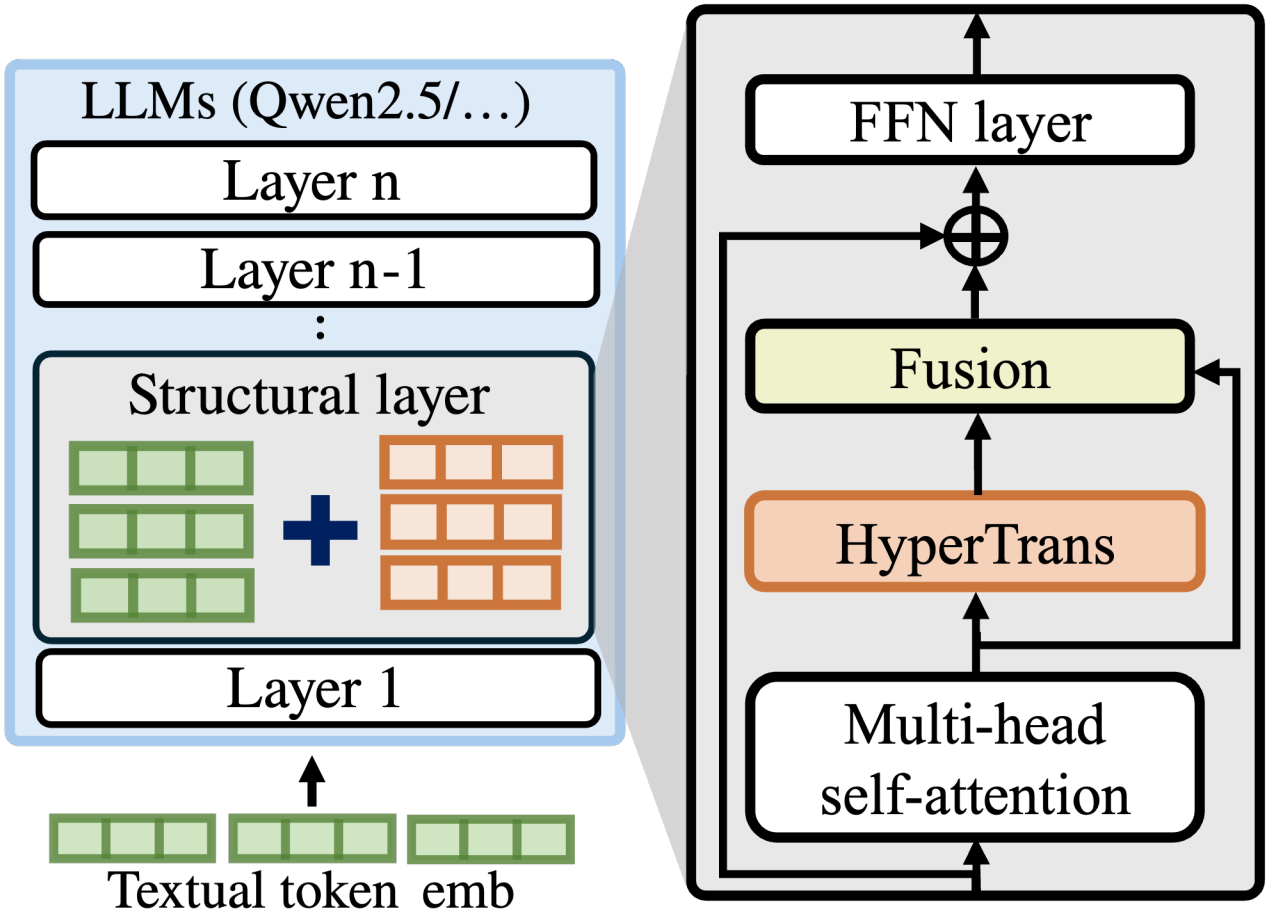
$$\mathbf{e}_v = \frac{1}{|\mathcal{T}_v^{\mathcal{E}}|} \sum_{i \in \mathcal{T}_v^{\mathcal{E}}} \mathbf{u}_i$$

$\mathcal{T}_u^{\mathcal{V}}, \mathcal{T}_v^{\mathcal{E}}$: The sets of tokens belonging to the u-th cell and v-th header

$\mathbf{u}_i$: Hidden representation of token i at the structural layer

Our approach: Injecting structural information at structural layers

Hypergraph message passing



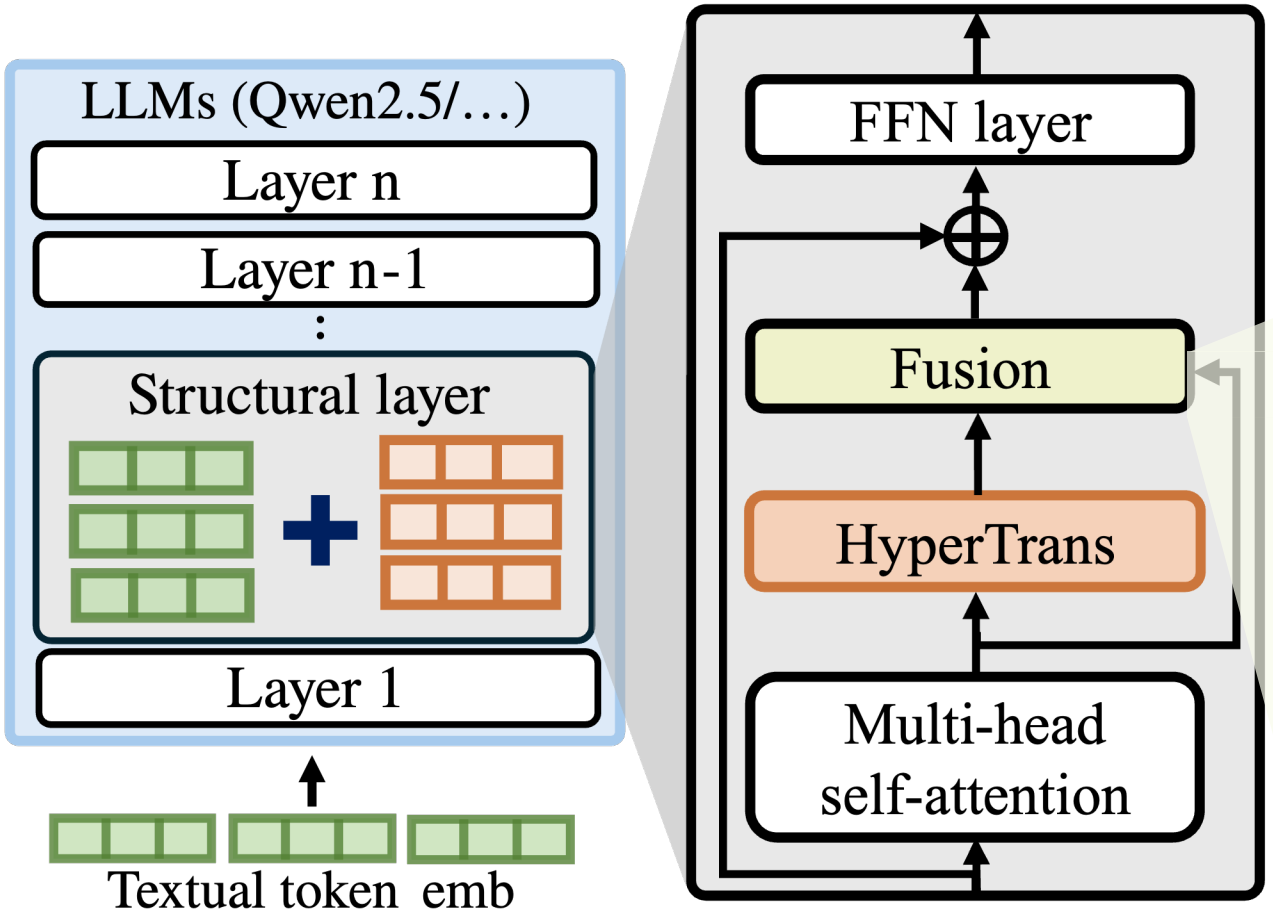
HyperTrans layer output

$$(\widetilde{\mathbf{V}}, \widetilde{\mathbf{E}}) = \mathrm{HyperTrans}(\mathbf{V}, \mathbf{E})$$

$$\widetilde{\mathbf{V}} = \{\widetilde{\mathbf{v}}_u\}_{u=1}^{N_{cell}} \quad \widetilde{\mathbf{E}} = \{\widetilde{\mathbf{e}}_v\}_{v=1}^{N_{header}}$$

Our approach: Injecting structural information at structural layers

Token-level fusion



HyperTrans layer output

$$(\widetilde{\mathbf{V}}, \widetilde{\mathbf{E}}) = \mathrm{HyperTrans}(\mathbf{V}, \mathbf{E})$$

$$\widetilde{\mathbf{V}} = \{\widetilde{\mathbf{v}}_u\}_{u=1}^{N_{cell}} \quad \widetilde{\mathbf{E}} = \{\widetilde{\mathbf{e}}_v\}_{v=1}^{N_{header}}$$

Structure feature

$$\widetilde{\mathbf{g}}_i = \begin{cases} \widetilde{\mathbf{v}}_u, & \text{if } i \in \mathcal{T}_u^{\mathcal{V}} \\ \widetilde{\mathbf{e}}_v, & \text{if } i \in \mathcal{T}_v^{\mathcal{E}} \end{cases}$$

Gating fusion

$$\alpha_i = \sigma\left(\mathbf{w}^\top [\, \mathbf{u}_i \,\|\, \widetilde{\mathbf{g}}_i \,]\right)$$

$$\widetilde{\mathbf{u}}_i = (1 - \alpha_i)\, \mathbf{u}_i + \alpha_i\, \widetilde{\mathbf{g}}_i.$$

$\alpha_i \in (0, 1)$: Sigmoid

$\mathcal{T}_u^{\mathcal{V}}, \mathcal{T}_v^{\mathcal{E}}$: The sets of tokens belonging to the u-th cell and v-th header

Results: Structural layers are the most effective injection points

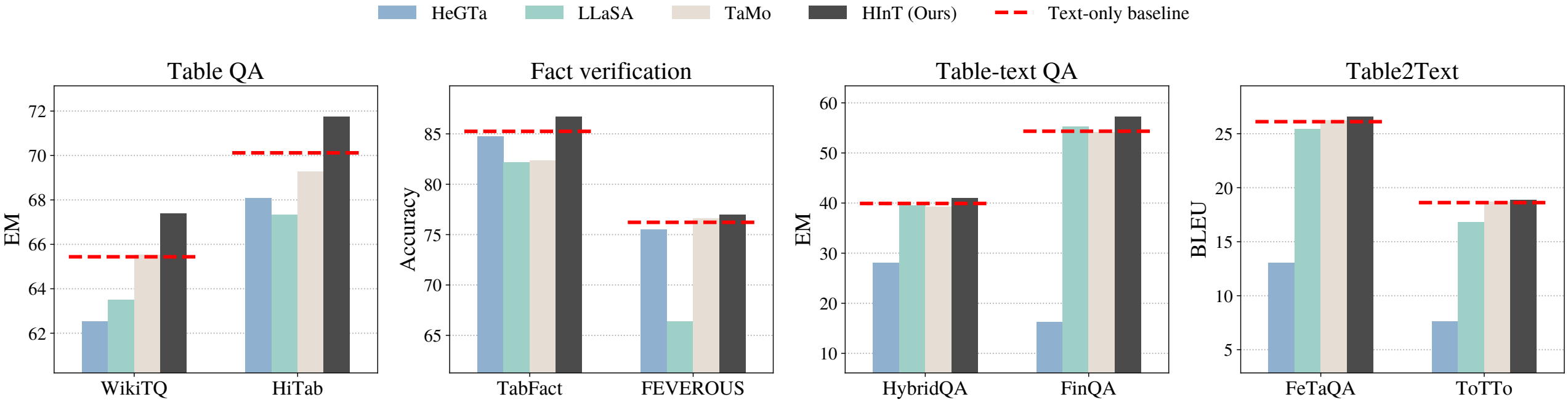
Consistently outperforms random or full-layer injection

Model: Qwen2.5-3B-Instruct

	Random	Full	**Structural Layers**
WikiTQ	65.71	66.37	**67.22**

(in % exact match; EM)

Results: HInT achieves the best overall performance



ICML
International Conference
On Machine Learning

MLV

Thank For Listening

POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY